

MedAlpaca -- An Open-Source Collection of Medical Conversational AI Models and Training Data

Year: 2023 | Citations: 373 | Authors: T. Han, L. Adams, Jens-Michalis Papaioannou, Paul Grundmann, Tom Oberhauser

Abstract

As large language models (LLMs) like OpenAI's GPT series continue to make strides, we witness the emergence of artificial intelligence applications in an ever-expanding range of fields. In medicine, these LLMs hold considerable promise for improving medical workflows, diagnostics, patient care, and education. Yet, there is an urgent need for open-source models that can be deployed on-premises to safeguard patient privacy. In our work, we present an innovative dataset consisting of over 160,000 entries, specifically crafted to fine-tune LLMs for effective medical applications. We investigate the impact of fine-tuning these datasets on publicly accessible pre-trained LLMs, and subsequently, we juxtapose the performance of pre-trained-only models against the fine-tuned models concerning the examinations that future medical doctors must pass to achieve certification.